

На правах рукописи  
УДК 81'33

**Серова Олеся Усмановна**

**ИНДИКАТОРЫ ФЕНОМЕНА ПЕРЕВОДЧЕСКОГО ЯЗЫКА В  
ПЕРЕВОДАХ НАУЧНЫХ СТАТЕЙ С АНГЛИЙСКОГО НА РУССКИЙ  
ЯЗЫК**

5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата филологических наук

Санкт-Петербург  
2026

Работа выполнена на кафедре образовательных технологий в филологии федерального государственного бюджетного образовательного учреждения высшего образования «Российский государственный педагогический университет им. А. И. Герцена»

**Научный руководитель:**

заслуженный деятель науки РФ, доктор филологических наук, профессор, профессор кафедры образовательных технологий в филологии федерального государственного бюджетного образовательного учреждения высшего образования «Российский государственный педагогический университет им. А. И. Герцена»

**Беляева Лариса Николаевна**

**Официальные оппоненты:**

доктор филологических наук, профессор, профессор департамента филологии, заведующий лабораторией языковой конвергенции федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики» – Санкт-Петербург»

**Колмогорова Анастасия Владимировна**

кандидат филологических наук, доцент кафедры английской филологии и перевода федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный экономический университет»

**Ефремова Мария Павловна**

**Ведущая организация:**

федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет»

Защита состоится 20 октября 2026 г. в 18 часов на заседании Совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук 33.2.018.08, созданного на базе Российского государственного педагогического университета им. А. И. Герцена, по адресу: 199004, г. Санкт-Петербург, 1-я линия В. О., д. 52, ауд. 48.

С диссертацией можно ознакомиться в фундаментальной библиотеке Российского государственного педагогического университета им. А. И. Герцена (191186, г. Санкт-Петербург, наб. р. Мойки, 48, корп. 5) и на сайте университета по адресу: [https://disser.herzen.spb.ru/Preview/Karta/karta\\_000001257.html](https://disser.herzen.spb.ru/Preview/Karta/karta_000001257.html).

Автореферат разослан «\_\_» \_\_\_\_\_ 2026 г.

Ученый секретарь диссертационного  
совета

Камшилова Ольга Николаевна

---

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Настоящее исследование посвящено изучению феномена переводческого языка на материале сопоставимых корпусов, включающих тексты непереводных русскоязычных научных статей и англо-русских ручных и машинных переводов.

Многочисленные исследования в области переводоведения демонстрируют статистически значимые отличия переводов от непереводных текстов, изначально созданных на языке перевода, по ряду лингвистических характеристик. Эти отличия дают основания говорить об особом, свойственном переводным текстам явлении – феномене переводческого языка (*translationese*). В переводе данный феномен проявляется через четыре основные характеристики: *упрощение, нормализацию, экспликацию и интерференцию*, каждая из которых представлена в тексте перевода соответствующими ей формальными индикаторами. Особенности переводных текстов, обусловленные этими характеристиками, нередко связывают со снижением качества перевода: считается, что переводы более низкого качества демонстрируют более выраженные признаки переводческого языка.

Анализ существующей литературы показал лакунарность работ, посвященных исследованию феномена переводческого языка на материале переводов научной литературы и в частности переводов научных статей с английского на русский язык. При этом в условиях стремительного роста публикационной активности ученых и расширения международного научного сотрудничества изучение данного феномена на материале переводов научных статей приобретает особую актуальность, поскольку связанное с ним снижение качества перевода способно приводить к неверной интерпретации данных исследования, а также подорвать доверие к автору и снизить влияние его работы в академическом сообществе. Установление индикаторов переводческого языка, потенциально влияющих на качество переводов научных статей, позволит не только лучше понять, как именно этот феномен проявляется в данной сфере перевода, но и выявить те аспекты переводов, которые необходимо учитывать как в самом процессе перевода, так и на этапе его постредактирования, в частности при постредактировании результатов машинного перевода (МП), что и определяет **актуальность** настоящей работы.

Кроме того, несмотря на активное использование корпусных методов в исследованиях феномена переводческого языка, на сегодняшний день не существует специализированных инструментов, позволяющих одновременно формировать сопоставимые корпуса текстов и проводить автоматизированный анализ индикаторов данного феномена на материале сформированных корпусов. Разработка такого инструмента имеет особую актуальность, поскольку систематическое сопоставление переводов с эталонными непереводными подкорпусами на основе формальных индикаторов переводческого языка позволит переводчикам и постредакторам количественно оценивать степень проявления этого феномена в конкретных текстах, выявляя наиболее проблемные аспекты перевода, требующие редактирования. Более того, в условиях растущей потребности в эффективных методах контроля качества переводов в академической сфере создание такого инструмента может способствовать

обеспечению более высокого качества переводов, что также определяет актуальность настоящего исследования.

**Степень научной разработанности темы.** Феномен переводческого языка активно исследуется с 1990-х годов. Теоретическую базу изучения данного феномена заложили труды М. Бейкер, Г. Тури, Э. Честермана, Ш. Блум-Кульки, С. Лавиозы, А. Мауранен и др. Корпусный подход к исследованию феномена был развит в работах С. Лавиозы, М. Бароне, С. Бернардینی, Ю. Илисей, В. Волански, Н. Ордана, Ш. Винтнера, Е. Лапшиновой-Колтунски, М. А. Куниловской и их соавторов. В отечественной традиции проблематика рассматривается в работах М. А. Куниловской, А. Б. Кутузова, Е. С. Краснопеевой, Н. Г. Кантышевой.

Вместе с тем до сих пор отсутствуют исследования феномена переводческого языка на материале англо-русских переводов научных статей, а также работы, предусматривающие сопоставление ручного и машинного перевода и переводов статей различных предметных областей в аспекте изучаемого явления.

**Теоретико-методологическую базу** исследования составили труды отечественных и зарубежных ученых, определяющие следующие аспекты темы:

- теория феномена переводческого языка и потенциальных универсальных характеристик переводов (Н. К. Гарбовский, Н. Г. Кантышева, М. А. Куниловская, Е. С. Краснопеева, М. Baker, М. Baroni, S. Bernardini, S. Blum-Kulka, A. Chesterman, I. Ilisei, M. Koppel, E. Lapshinova-Koltunski, S. Laviosa, A. Mauranen, N. Ordan, V. Volansky, S. Wintner);

- корпусные исследования в переводоведении (В. П. Захаров, М. Baker, S. Laviosa, M. Snell-Hornby, A. Chesterman, S. Granger, T. McEnery, A. Hardie);

- теория научного перевода и сопоставление особенностей научного стиля в русском и английском языках (И. С. Алексеева, И. В. Арнольд, Л. Н. Беляева, Р. К. Боженкова, А. Н. Васильева, И. Р. Гальперин, В. В. Гуревич, О. Н. Камшилова, М. Н. Кожина, В. Н. Комиссаров, Т. П. Плещенко, А. Л. Пумпянский, А. В. Федоров);

- методология корпусных исследований в аспекте изучения феномена переводческого языка (Е. С. Краснопеева, М. А. Куниловская, А. Б. Кутузов, М. Baker, M. Baroni, S. Bernardini, S. Blum-Kulka, G. Corpas Pastor, I. Ilisei, M. Koppel, E. Lapshinova-Koltunski, S. Laviosa, N. Ordan, V. Volansky, S. Wintner).

**Объект** исследования – перевод научных статей с английского языка на русский. **Предмет** исследования – индикаторы феномена переводческого языка, характерные для переводов научных статей с английского на русский язык.

**Гипотеза исследования** состоит в том, что переводы научных статей с английского языка на русский демонстрируют статистически значимые отличия от непереводаемых русскоязычных научных статей по ряду признаков, соответствующих четырем основным характеристикам переводческого языка (*упрощению, нормализации, экспликации и интерференции*), причём часть этих отличий, связанная с отклонениями от норм русского научного стиля, может служить потенциальными предикторами качества переводного текста.

**Цель работы** – установление индикаторов феномена переводческого языка, влияющих на качество переводов научных статей с английского на русский язык.

Для достижения поставленной цели и проверки выдвинутой гипотезы были решены следующие задачи:

1) Изучить исторические предпосылки, обусловившие интерес к исследованиям феномена переводческого языка, а также существующие подходы к его определению;

2) Рассмотреть концептуальные подходы к характеристикам переводческого языка как к переводческим универсалиям и как к потенциально универсальным характеристикам переводных текстов, а также существующие подходы к исследованию этих характеристик;

3) Описать особенности научного стиля в английском и русском языках для формирования теоретической базы для интерпретации результатов корпусного анализа;

4) Отобрать формальные индикаторы потенциально универсальных характеристик переводов, а также индикаторы, отражающие нормы научного стиля русского языка, с целью их последующего анализа;

5) Разработать программный комплекс для автоматизированного анализа формальных индикаторов на материале исследовательских текстовых корпусов;

6) Сформировать два сопоставимых исследовательских корпуса: корпус научных статей социологической тематики (непереводные тексты и ручные переводы) и корпус научных статей медицинской тематики (непереводные тексты, ручные и машинные переводы), проанализировать сформированные корпуса при помощи разработанного программного комплекса и провести статистический анализ полученных результатов с применением методов оценки значимости различий между подкорпусами;

7) Основываясь на результатах анализа, определить набор индикаторов, которые могут быть использованы при оценке качества переводов научных статей с английского на русский язык, и сформулировать практические рекомендации для переводчиков и постредакторов.

**Материалом исследования** послужили тексты научных статей, отобранные из рецензируемых научных периодических изданий, сгруппированные по тематике и способу перевода и сгруппированные в три сопоставимых корпуса:

1. Корпус статей социологической тематики: 59 непереводных статей и 59 ручных переводов (по  $\approx 310\ 000$  словоформ в каждом подкорпусе), извлечённых из журналов «Социологическое обозрение», «Экономическая социология», «Журнал социологии и социальной антропологии»;

2. Корпус статей медицинской тематики: 70 непереводных статей и 70 ручных переводов (по  $\approx 216\ 500$  словоформ в подкорпусе), извлечённых из рецензируемых медицинских журналов («Нефрология и диализ», «Рациональная Фармакотерапия в Кардиология», «Доказательная Гастроэнтерология», «The EYE ГЛАЗ», «Российский кардиологический журнал», «Патология кровообращения и кардиохирургия», «Остеопороз и остеопатии», «Кардиология», «Качественная клиническая практика», «Нервно-мышечные болезни», «Русский журнал детской неврологии», «Пульмонология», «Проблемы эндокринологии», «Клиническая и экспериментальная тиреоидология», «Проблемы репродукции»);

3. Корпус статей медицинской тематики: те же 70 непереводаемых медицинских статей, а также 50 машинных переводов, выполненных системой Google Translate, оригиналы которых взяты из англоязычных журналов «The Lancet», «The New England Journal of Medicine», «Biology of Reproduction», «European Journal of Pharmacology» (по  $\approx 216\ 500$  словоформ в подкорпусе).

Основными методами исследования послужили:

(1) гипотетико-дедуктивный и индуктивный методы для формулирования исходной гипотезы о наличии характеристик переводческого языка в переводах научных статей и для обобщения полученных эмпирических данных исследования;

(2) метод лингвистического эксперимента для сопоставления переводных и непереводаемых подкорпусов в контролируемых условиях (единые критерии отбора текстов по жанру, тематике, периоду создания) для проверки выдвинутой гипотезы;

(3) метод корпусного анализа, реализованный при помощи разработанного в рамках исследования программного комплекса для анализа индикаторов феномена переводческого языка на материале корпусов непереводаемых русскоязычных текстов и англо-русских переводов;

(4) методы статистического анализа для выявления статистически значимых различий между подкорпусами: двухвыборочный  $t$ -критерий Уэлча для показателей, продемонстрировавших распределение данных, близкое к нормальному, а также непараметрический  $U$ -критерий Манна–Уитни для анализа показателей, чье распределение оказалось отличным от нормального. Для оценки практической значимости различий рассчитывалась величина эффекта  $r$ .

**На защиту выносятся следующие положения:**

1) переводы научных статей с английского языка на русский демонстрируют статистически значимые отличия от непереводаемых русскоязычных статей по ряду индикаторов, соответствующих основным характеристикам переводческого языка (*упрощению, нормализации, экспликации и интерференции*);

2) каждая из указанных характеристик (*упрощение, нормализация, экспликация и интерференция*) проявляется в переводах научных статей через совокупность поддающихся количественному анализу формальных индикаторов;

3) разработанный в рамках исследования программный комплекс позволяет автоматизированно выявлять формальные индикаторы переводческого языка при сопоставлении подкорпуса переводных текстов с эталонным подкорпусом непереводаемых текстов на русском языке;

4) степень выраженности индикаторов переводческого языка варьируется как в зависимости от предметной области переводных научных статей (социологическая / медицинская тематика), так и от способа выполнения перевода (ручной / машинный), причём машинные переводы демонстрируют более выраженные отклонения от норм русского научного стиля, чем ручные переводы;

5) установленные индикаторы переводческого языка могут быть разделены на две группы: (а) индикаторы, фиксирующие отклонения от норм русского научного стиля, потенциально снижающие качество перевода и требующие учёта при переводе и постредактировании, и (б) индикаторы, фиксирующие отличия переводных текстов от непереводаемых, но либо не связанные

со снижением качества перевода напрямую, либо требующие дальнейших исследований для установления характера их влияния на качество перевода.

**Научная новизна исследования** заключается в том, что впервые:

1) проведён комплексный анализ феномена переводческого языка на материале переводов научных статей с английского языка на русский с использованием 46 формальных индикаторов данного феномена, охватывающих четыре потенциально универсальные характеристики переводов (*упрощение, нормализацию, экспликацию и интерференцию*), а также индикаторы соответствия нормам научного стиля русского языка;

2) разработан специализированный программный комплекс для автоматизированного сбора корпуса текстов и дальнейшего анализа сформированного корпуса с целью установления формальных индикаторов феномена переводческого языка в переводных текстах;

3) на основе сопоставительного анализа непереводных текстов с текстами ручных и машинных переводов двух предметных областей (социологической и медицинской) произведена дифференциация проявлений переводческого языка на (а) потенциально влияющие на качество переводов научных статей и (б) фиксирующие отличия переводных текстов от непереводных, но не связанные напрямую со снижением качества перевода.

**Теоретическая значимость** исследования заключается в следующем:

1) получили развитие теоретические представления о проявлении потенциально универсальных характеристик перевода на конкретной языковой паре (английский → русский) и жанре (научная статья);

2) разработанная методология эмпирического исследования вносит вклад в развитие представлений о проектировании и использовании специализированных программных инструментов для корпусных исследований феномена переводческого языка.

**Практическая значимость** работы определяется:

1) возможностью применения полученных результатов в таких практических сферах, как перевод научных статей, постредактирование ручных и машинных переводов, разработка критериев оценки качества ручных и машинных переводов и подготовка переводчиков научной литературы;

2) возможностью использования разработанного программного комплекса переводчиками и постредакторами с целью выявления индикаторов переводческого языка на материале выполняемых / редактируемых переводов, а также с целью сопоставления результатов анализа этих переводов с результатами анализа непереводных текстов (при наличии эталонного непереводного подкорпуса), что может способствовать более объективной оценке качества ручных и машинных переводов и выявлению областей, требующих улучшения.

**Степень достоверности результатов** обеспечивается объёмом проанализированного корпусного материала (308 текстов статей социологической / медицинской тематики общим объёмом около 1,27 млн словоформ), применением методов корпусного анализа, а также статистического анализа (двухвыборочный t-критерий Уэлча и непараметрический U-критерий Манна–Уитни).

**Апробация результатов исследования.** Основные положения и результаты исследования были изложены в докладах, представленных на научных и научно-практических конференциях российского и международного уровней:

1. Международная научная конференция «Корпусная лингвистика – 2023» (Санкт-Петербург, 21–23 июня 2023 г., СПбГУ);
2. II Международная научная филологическая конференция имени Л. А. Вербицкой (Санкт-Петербург, 19–26 марта 2024 г., СПбГУ);
3. IV Международная конференция «Пиотровские Чтения – 2024» (Санкт-Петербург, 22 ноября 2024 г., РГПУ им. А. И. Герцена).
4. Всероссийская научно-практическая конференция с международным участием «Язык, текст, культура через призму цифровых технологий» (Санкт-Петербург, 16-17 декабря 2024 г., ВШЭ).
5. II Всероссийская научно-практическая конференция с международным участием «Язык, текст, культура через призму цифровых технологий» (Санкт-Петербург, 4-5 декабря 2025 г., ВШЭ).

**Структура и объем диссертационной работы.** Диссертация состоит из введения, трех глав, заключения, списка сокращений и условных обозначений, списка использованной литературы, включающего 96 наименований на русском языке, 127 наименований на английском языке, списка источников, из которых были взяты статьи для формирования исследовательских корпусов, и 9 приложений. Общий объем диссертации составляет 220 страниц печатного текста, из них 188 страниц занимает основное содержание работы, включая 22 таблицы, и 32 страницы содержат приложения.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

Во **введении** обосновываются выбор темы и её актуальность, определяются цель, задачи, объект, предмет и гипотеза исследования, описывается используемая теоретико-методологическая база, материал и методы, формулируется теоретическая и практическая значимость и новизна исследования, а также положения, выносимые на защиту, описывается апробация результатов, а также структура, объем и содержание исследования.

**Глава 1 «Предпосылки и теоретические основы исследования феномена переводческого языка»** состоит из 4 разделов.

В **разделе 1.1.** рассмотрены исторические предпосылки исследований феномена переводческого языка. Установлено, что научный интерес к данному феномену сформировался под влиянием теоретико-методологических сдвигов, произошедших в переводоведении в последней четверти XX века: методологический поворот от прескриптивного переводоведения к дескриптивному, возросшее понимание важности переводной литературы как одного из главных средств межкультурного трансфера знаний, а также признание переводного дискурса специфическим элементом дискурсивного пространства принимающей культуры подготовили почву для того, чтобы перевод стал рассматриваться как самостоятельный объект научных исследований, а изучение характерных особенностей переводных текстов приобрело системный характер.

**В разделе 1.2.** дан обзор существующих подходов к пониманию феномена переводческого языка. Специфические особенности переводных текстов принято рассматривать с разных позиций: (1) как результат интерференции исходного языка и текста, (2) как результат сложного взаимодействия системы языка оригинала и языка перевода, а также (3) как результат, обусловленный самой природой переводческого процесса, вне зависимости от задействованной в переводе языковой пары. Кроме того, наблюдается неоднородность в оценке влияния данного феномена на качество перевода. Она варьируется от негативной – когда переводческий язык рассматривается как явление, свидетельствующее о недостаточном внимании переводчика к специфике переводящего языка (ПЯ) и чрезмерном стремлении к сохранению формальных структур оригинала, – до нейтральной, в рамках которой феномен трактуется как закономерное, в некотором смысле естественное явление, систематически проявляющееся в переводах вследствие объективных различий в культурных и жанрово-стилистических традициях исходного и переводящего языков. Несмотря на разнообразие трактовок, все они признают объективное существование специфических характеристик, систематически проявляющихся в переводах и отличающих их от непереводных текстов на ПЯ, что обуславливает необходимость эмпирического исследования этих характеристик на конкретном языковом материале.

**В разделе 1.3.** рассматривается концепция переводческих универсалий М. Бейкер, ставшая отправной точкой для современных исследований феномена переводческого языка и поставившая вопрос об универсальности его характеристик для всех переводных текстов вообще. Рассматривается существующая критика концепции переводческих универсалий, согласно которой утверждать абсолютную универсальность выявляемых в переводе характеристик некорректно ввиду зависимости их проявления от разнообразных переменных (языковой пары, жанрово-стилистической специфики текста и т. д.), в связи с чем имеет смысл говорить о переводческих характеристиках лишь как об исследовательских гипотезах, которые нуждаются в строгой эмпирической проверке на конкретном языковом материале, ограниченном четко определенными критериями отбора. Кроме того, в разделе 1.3. рассматриваются два подхода к исследованию переводческих характеристик: (1) подход «сверху вниз» (от формулирования гипотезы об универсальности характеристик, её проверки на максимально большом и разнообразном текстовом материале к ограничению и уточнению изначальной гипотезы) и подход «снизу вверх» (от формулирования частного утверждения о наличии переводческих характеристик для ограниченного определёнными критериями подмножества текстов, проверки этого утверждения к постепенному расширению области применения изначального утверждения через расширение спектра анализируемого материала). В настоящей работе за основу был принят подход «снизу вверх», поскольку он позволяет избежать преждевременных обобщений об универсальности изучаемого феномена и обеспечивает более высокую степень достоверности делаемых выводов.

**В разделе 1.4.** описываются стилеобразующие особенности научного стиля в русском и английском языках. Их рассмотрение сыграло ключевую роль как при отборе дополнительной группы индикаторов для анализа исследуемого феномена,

так и при интерпретации результатов корпусного анализа. Показано, что русскоязычный научный текст характеризуется более высокой степенью номинативности, в нем преобладают существительные в форме единственного числа (для обозначения общего понятия или целого класса объектов), выражена тенденция к субстантивации глаголов, среди степеней сравнения предпочтительны аналитические формы, активно используется страдательный залог (аналитические конструкции и возвратные конструкции на *-ся*), преобладают глаголы несовершенного вида, глаголы настоящего времени во вневременном значении, нередко используются номинативные рамочные конструкции с препозитивным определением, личные местоимения первого лица практически не употребляются и т. д. Англоязычный научный текст, напротив, отличается большей глагольностью, менее выраженной тенденцией к субстантивации, более широким использованием существительных во множественном числе (в т.ч. для обозначения целого класса объектов), а также активным использованием личных местоимений первого лица в связи с характерной для современного англоязычного научного дискурса тенденцией к подчёркиванию личного вклада исследователя и т. д.

**Глава 2 «Отбор индикаторов для исследования характеристик феномена переводческого языка»** состоит из 4 разделов и посвящена подробному описанию четырех основных характеристик переводческого языка, а также 46 формальных индикаторов, отобранных для дальнейшего корпусного анализа.

**Раздел 2.1.** посвящен характеристике *упрощение*, которая основывается на предположении о том, что переводчики склонны делать текст перевода лексически и грамматически более простым вследствие стремления избежать потенциальной неверной интерпретации текста читателем. Для анализа отобрано 16 формальных индикаторов: (1-4) коэффициент лексической вариативности (базовая версия – *TTR*, а также версии, нивелирующие зависимость показателя от длины текста: *Log TTR*, *Modified TTR*, *Standardized TTR*), (5) коэффициент лексической плотности, отражающий соотношение содержательных слов к общему объёму текста в токенах, (6) средний ранг используемой лексики, вычисляемый на основе ранжирования слов текста в частотном словаре русской лексики, (7) средняя доля наиболее частотных знаменательных слов подкорпуса в тексте, (8-9) средняя длина слов (символы и слоги), (10-12) предложений (символы, токены, клаузы) и (13) клауз (токены), (14) средняя доля простых предложений, (15) индекс удобочитаемости Флэша и (16) нормализованные частоты знаков препинания.

**В разделе 2.2.** рассматривается характеристика *нормализация*, отражающая тенденцию к употреблению в переводе более типизированных форм и структур ПЯ и к избеганию лексических повторов. Для анализа выбрано 3 индикатора: (1) лексическая повторяемость знаменательной лексики, (2) среднее значение показателя поточечной взаимной информации (*Pointwise Mutual Information*) словарных биграмм в 2 вариациях (*PMI* и *Modified PMI*), (3) доля биграмм, значение *PMI* / *Modified PMI* которых выше 0. Под поточечной взаимной информацией понимается показатель, отражающий силу ассоциативной связи между двумя соседними словами, то есть насколько чаще два слова встречаются совместно, чем можно было бы ожидать, если бы их появления были независимы друг от друга.

Более высокое среднее значение показателя говорит об использовании в подкорпусе более устойчивых, предсказуемых коллокаций ПЯ.

В разделе 2.3. рассматривается характеристика *экспликация*, проявляющаяся в повышении объяснительного потенциала перевода за счет использования средств обеспечения лексической и грамматической когезии. Для анализа лексической когезии выбраны индикаторы: среднее косинусное сходство (1) пар слов и (2) соседних предложений в тексте – показатели степени семантической близости лексики, используемой (1) во всем тексте и (2) в идущих друг за другом предложениях. Для пар слов более высокое среднее значение косинусного сходства будет говорить о наличии в тексте более связанных по смыслу слов (сино-, антонимов, гиперо- и гипонимов, холо- и меронимов, коллокаций), а для соседних предложений – о более сильной семантической связности между предложениями в тексте и, следовательно, о большей плавности смысловых переходов между ними. Для анализа грамматической когезии выбраны: (3) эксплицитное название – отношение числа личных и притяжательных местоимений к общему числу именованных сущностей в тексте, нормализованные частоты дискурсивных маркеров (6) локальной и (7) глобальной связности текста. Дополнительно анализировались: (4) единичное именование – доля именованных сущностей, состоящих из одного токена, и (5) средняя длина именованных сущностей.

В разделе 2.4. рассматривается характеристика *интерференция*, состоящая в переносе в текст перевода структурных особенностей исходного языка и текста. Для анализа характеристики выбран 1 индикатор – нормализованные частоты частеречных n-грамм (уни-, би- и триграмм). Для частеречной разметки текстов использовались библиотеки *Natasha* и *rumorphy2*. Для более детализированного исследования анализ был расширен системой специальных тегов, позволяющих фиксировать разнообразные синтаксические структуры: придаточные предложения с союзами и союзными словами, аналитические и синтетические формы степеней сравнения, составные именные и глагольные сказуемые, конструкции с модальными глаголами и словами, вводные конструкции, конструкции с сослагательным наклонением, деепричастные и причастные обороты, конструкции с причастием в функции препозитивного определения.

В разделе 2.5. описаны индикаторы, отобранные для установления соответствия переводных текстов нормам русского научного стиля. В группу вошло 19 индикаторов: нормализованные частоты дискурсивных маркеров (1) авторского начала в тексте и (2) взаимодействия автора с читателем; (3) доля существительных в тексте; доля существительных (4) среднего рода, (5) в форме единственного числа; (6) доля абстрактных существительных; (7) доля прилагательных в тексте, (8) доля кратких прилагательных, доля прилагательных в формах (9-10) сравнительной и превосходной степени; (11) доля глаголов в тексте, (12) доля глаголов в форме страдательного залога, (13) доля причастий относительно всех глаголов, (14) доля причастий в препозиции перед существительным, (15) доля причастий в составе причастных оборотов, (16) доля деепричастий относительно всех глаголов, (17) доля глаголов несовершенного вида, (18) доля глаголов в форме прошедшего / настоящего / будущего времени и (19) нормализованные частоты местоимений разных разрядов.

**Глава 3 «Исследование характеристик переводческого языка на материале научных статей»** посвящена формированию исследовательских корпусов, разработке программного комплекса для анализа отобранных индикаторов, описанию и обоснованию выбора методов статистической оценки значимости различий между подкорпусами, а также интерпретации результатов.

В разделе 3.1. описывается процедура формирования и подготовки к анализу трёх сопоставимых корпусов научных статей: (1) социологического (непереводные тексты и ручные переводы); (2) медицинского (непереводные тексты и ручные переводы); (3) медицинского (непереводные тексты и машинные переводы). Описываются критерии отбора исследовательского материала: соответствие текстов единому жанру (научная статья), тематике (социология или медицина) и периоду создания (1990-2025 гг.). Также описывается процедура предобработки анализируемого материала: удаление аннотаций, ключевых слов, таблиц, сносок, списков литературы вручную, а также токенизация, лемматизация, морфологическая разметка текстов с помощью библиотек *Natasha* и *py morphology2*. Кроме того, в рамках раздела на базе языка Python был разработан специализированный программный комплекс – анализатор феномена переводческого языка (<https://pypi.org/project/translationese-analyzer/>), – обеспечивший автоматизированный сбор исследовательских корпусов, их предобработку, вычисление показателей 46 индикаторов и сопоставление результатов анализа между переводными и непереводными подкорпусами.

В разделе 3.2. излагается методология статистического анализа полученных данных. Для показателей с распределением, близким к нормальному, применялся двухвыборочный *t*-критерий Уэлча, адаптированный для корпусных данных с неравными объёмами текстов (каждый текст получал вес, пропорциональный числу содержащихся в нём словарных токенов). На основе взвешенных средних и взвешенных дисперсий вычислялась *t*-статистика; эффективный размер выборки оценивался по формуле Киша. При сопоставлении подкорпусов фиксировались: разность взвешенных средних ( $\Delta$ ), *p*-значение, 95%-й доверительный интервал для  $\Delta$ . Дополнительно для каждого подкорпуса вычислялся коэффициент вариации (*CV*) как мера внутригрупповой колеблемости данных. Для показателей индикаторов с распределением, отличным от нормального (индикаторов, связанных с употреблением именованных существностей, и дискурсивных маркеров), применялся непараметрический *U*-критерий Манна–Уитни. В данном случае, наряду с *p*-значением, вычислялись: оценка сдвига Ходжеса–Лемана (*HL*) – медиана всех попарных разностей между наблюдениями двух групп, служащая робастным аналогом  $\Delta$ , а также эффект  $A_{12}$  (вероятность того, что случайно выбранное значение из непереводного подкорпуса превысит значение из переводного). В обоих случаях статистическая значимость различий оценивалась на уровне  $\alpha=0.05$  (в таблицах с результатами анализа ниже *n.s.* соответствует  $p \geq 0.05$ , \* –  $p < 0.05$ , \*\* –  $p < 0.01$ , \*\*\* –  $p < 0.001$ ), а практическая значимость – через величину эффекта *r* ( $r \approx 0.1$  – малый эффект,  $r \approx 0.3$  – средний,  $r \geq 0.5$  – большой).

В разделе 3.3. представлена интерпретация результатов анализа индикаторов. Далее для удобства изложения результатов используются следующие обозначения: СОЦ-РП, когда речь идет о ручных переводах по социологии; МЕД-РП, если речь

о ручных переводах по медицине; МЕД-МП – если речь о машинных переводах по медицине. В приводимых далее таблицах  $\Delta$  обозначает разность взвешенных средних между непереводам и переводным подкорпусами, *HL* – оценку сдвига Ходжеса–Лемана. Оба показателя служат мерой величины и направления различий между подкорпусами. Положительное значение  $\Delta / HL$  означает, что показатель в переводном подкорпусе ниже, чем в непереводам; отрицательное – что выше. Начнем с характеристики *упрощение* (Табл. 1).

Таблица 1 – Индикаторы характеристики *упрощение*

Индикатор	СОЦ-РП	МЕД-РП	МЕД-МП
Лексическая плотность	$\Delta=+2.25$ ***, $r\approx 0.4$	$\Delta=+1.91$ ***, $r\approx 0.31$	$\Delta=+4.01$ ***, $r\approx 0.54$
TTR	n.s.	$\Delta=+2.85$ **, $r\approx 0.26$	$\Delta=+5.11$ ***, $r\approx 0.53$
Log TTR	n.s.	$\Delta=+1.28$ **, $r\approx 0.30$	$\Delta=+1.81$ ***, $r\approx 0.47$
Modified TTR	n.s.	$\Delta=+84.27$ **, $r\approx 0.27$	n.s.
STTR	n.s.	$\Delta=+2.75$ **, $r\approx 0.28$	$\Delta=+4.62$ ***, $r\approx 0.49$
Средний ранг слов (MWR1)	$\Delta=+138$ ***, $r\approx 0.4$	$\Delta=+273.9$ ***, $r\approx 0.46$	n.s.
Средний ранг слов (MWR2)	$\Delta=+29.18$ *, $r\approx 0.23$	n.s.	$\Delta=+53.07$ **, $r\approx 0.31$
Средняя длина слов (в символах)	$\Delta=+0.2$ ***, $r\approx 0.32$	n.s.	$\Delta=+0.31$ ***, $r\approx 0.46$
Средняя длина слов (в слогах)	$\Delta=+0.08$ **, $r\approx 0.31$	n.s.	$\Delta=+0.06$ *, $r\approx 0.22$
Средняя длина предложений (в символах)	$\Delta=+29.5$ ***, $r\approx 0.46$	n.s.	$\Delta=-11.87$ *, $r\approx 0.26$
Средняя длина предложений (в токенах)	$\Delta=+4.8$ ***, $r\approx 0.44$	n.s.	$\Delta=-3.63$ **, $r\approx 0.43$
Средняя длина предложений (в клаузах)	n.s.	$\Delta=-0.13$ **, $r\approx 0.28$	$\Delta=-0.20$ ***, $r\approx 0.47$
Средняя длина клауз (в токенах)	$\Delta=+1.0$ ***, $r\approx 0.43$	n.s.	n.s.
Доля простых предложений	n.s.	$\Delta=+4.42$ **, $r\approx 0.26$	$\Delta=+8.15$ ***, $r\approx 0.49$
Индекс удобочитаемости	$\Delta=-9.4$ ***, $r\approx 0.44$	n.s.	n.s.

В СОЦ-РП, МЕД-РП и в особенности в МЕД-МП снижается *лексическая плотность*: переводы содержат больше функциональной и меньше знаменательной лексики в сравнении с непереводам текстами. *Лексическая вариативность* снижается только в подкорпусах медицинских переводов (МЕД-РП и МЕД-МП), причем в МЕД-МП эффект выражен в значительно большей степени. Примечательно, что модифицированная версия показателя (*Modified TTR*) снижается только в ручных переводах (МЕД-РП), что говорит о снижении в них доли лексем, встречающихся в тексте единожды (*hapax legomena*).

Во всех переводных подкорпусах прослеживается тенденция к использованию более частотной лексики ПЯ. В СОЦ-РП снижение среднего ранга слов фиксируется по обеим версиям индикатора (*MWR1* и *MWR2*), что говорит об общей тенденции к выбору более частотных и, следовательно, с большей вероятностью знакомых читателю единиц ПЯ. В МЕД-РП различия выявлены только по *MWR1*: разница между переводами и непереводам текстами сосредоточена именно в зоне сверхредкой лексики (с рангом более 6000), куда входит большая часть медицинской терминологии. В МЕД-РП различия выявлены только по *MWR1* – это

означает, что в целом система МП справляется с передачей редкой лексики, однако в пределах первых 6000 единиц списка тяготеет к выбору более высокочастотных вариантов. Наблюдаемая закономерность требует особого внимания при переводе и постредактировании, поскольку в случае замены низкочастотных терминов более общеупотребительными вариантами ПЯ может снижаться точность перевода. *Средняя длина слов* снижается в СОЦ-РП и МЕД-МП. Причем в МЕД-МП результат может объясняться тем, что современные системы МП тяготеют к выбору более высокочастотных соответствий ПЯ, которые в среднем короче более редких лексем. Показатели *средней длины предложений* и *клауз* демонстрируют разнонаправленные тенденции. В СОЦ-РП предложения (в токенах и символах) и клаузы (в токенах) оказались короче при неизменном числе клауз на предложение, что говорит о реализации *упрощения* в переводе не через изменение структурной сложности предложений (число клауз сохраняется), а через использования более компактных словосочетаний внутри клаузы. В МЕД-РП *средняя длина предложений* (в токенах и символах) не меняется, однако растет *длина предложений в клаузах* и снижается *доля простых предложений* – результат, который может указывать на *экспликацию* логических связей оригинала (например, через более частое употребление СПП и ССП). В МЕД-МП предложения становятся длиннее (в токенах, символах и клаузах), а *доля простых предложений* резко снижается, что, вероятно, свидетельствует о калькировании системой МП синтаксических структур оригинала и говорит о её неспособности адекватно реализовывать необходимые синтаксические перестройки. При постредактировании результатов МП следует это учитывать, разбивая структурно переусложнённые предложения на более простые. *Индекс удобочитаемости* растет только в СОЦ-РП, что указывает на бóльшую лёгкость восприятия переводных текстов по социологии по сравнению с непереводаемыми.

Рассмотрим результаты анализа пунктуационных знаков (Табл. 2).

Таблица 2 – Индикаторы характеристики *упрощение* (пунктуация)

Знак	СОЦ-РП	МЕД-РП	МЕД-МП
.	$\Delta=-0.45$ ***, $r\approx 0.35$	n.s.	$\Delta=+0.39$ ***, $r\approx 0.34$
–	$\Delta=+0.16$ *, $r\approx 0.25$	$\Delta=+0.28$ ***, $r\approx 0.44$	$\Delta=+0.48$ ***, $r\approx 0.79$
(	$\Delta=+0.19$ *, $r\approx 0.19$	$\Delta=+0.53$ ***, $r\approx 0.38$	$\Delta=-0.34$ ***, $r\approx 0.25$
)	$\Delta=+0.21$ *, $r\approx 0.22$	$\Delta=+0.52$ ***, $r\approx 0.37$	$\Delta=-0.31$ *, $r\approx 0.24$
« и »	$\Delta=+0.56$ ***, $r\approx 0.39$	n.s.	n.s.
,	n.s.	n.s.	$\Delta=-0.64$ ***, $r\approx 0.31$

Частота использования *точек* растёт в СОЦ-РП, что в целом согласуется с сокращением *средней длины предложений*: переводчики дробят более длинные предложения оригинала на более короткие в переводе. В МЕД-МП частота *точек*, напротив, снижается, что может свидетельствовать в пользу уже упомянутой тенденции к воспроизведению системами МП синтаксической структуры оригинала. Во всех трёх переводных подкорпусах фиксируется снижение частоты *тире*, причём в МЕД-МП эффект наиболее выраженный ( $r\approx 0.79$ ). Результат обусловлен функциональной асимметрией данного знака в русском и английском языках: в русском тексте *тире* выполняет функции, не имеющие прямого

соответствия в английском (замена глагола-связки, всего сказуемого, выражение логических отношений при бессоюзной связи). Поскольку специфические для русского языка случаи эллипсиса, требующие постановки *there*, исходным текстам не свойственны, то и в переводе *there* встречаются значительно реже. В СОЦ-РП и МЕД-РП круглые скобки используются реже, чем в непереводах, что, вероятно, отражает тенденцию к редукции парентетической сложности (круглые скобки часто используются для введения дополнительной информации, уточнений и пояснений): вероятно, переводчики стремятся сделать изложение более линейным, интегрируя представленную в скобках информацию в основной текст. В то же время в МЕД-МП наблюдается противоположная тенденция, что говорит о переносе системой МП структур оригинала в текст перевода без интеграции содержимого скобок в основной текст. Снижение частоты кавычек-ёлочек именно в СОЦ-РП закономерно: социологические статьи нередко включают цитаты, фрагменты прямой речи. А полученный результат предположительно обусловлен двумя причинами: тем, что в английском языке принято использовать иные типы кавычек (" "), и тем, что прямая речь оригинала может передаваться в переводе как косвенная, интегрируясь в основной текст. Наконец, рост частоты запятых в МЕД-МП согласуется с увеличением *числа клауз на предложение* и подтверждает перенос синтаксической структуры исходных предложений в текст МП.

Рассмотрим результаты анализа характеристики *нормализация*. Показатель *лексической повторяемости* снижается в СОЦ-РП ( $\Delta=+0.35^*$ ,  $r\approx 0.21$ ) и в особенности в МЕД-МП ( $\Delta=+1.76$ ,  $***$ ,  $r\approx 0.17$ ). В случае ручных переводов по социологии это снижение можно интерпретировать как проявление *нормализации*: переводчики стремятся избегать повторного использования одних и тех же знаменательных лексем на протяжении текста. В случае с МП снижение обусловлено иным механизмом: системы МП могут передавать одну и ту же лексему оригинала разными способами из-за перевода по предложениям, что приводит к нарушению лексической (а в случае с переводами научных статей – терминологической) консистентности текста и что важно учитывать при постредактировании результатов МП, целенаправленно проверяя единообразие передачи терминологической лексики.

Анализ индикаторов *поточечной взаимной информации* показал, что в ручных переводах используются коллокации, являющиеся менее устойчивыми словосочетаниями ПЯ, что противоречит гипотезе о *нормализации*. В МП, напротив, употребляются более предсказуемые, шаблонные словосочетания ПЯ.

Рассмотрим результаты анализа характеристики *экспликация* (Табл. 3).

Таблица 3 – Индикаторы характеристики *экспликация*

Индикатор	СОЦ-РП	МЕД-РП	МЕД-МП
Единичное именование	n.s.	HL=-16.67 **, $r\approx 0.27$	HL=-21.46 ***, $r\approx 0.47$
Средняя длина именованных сущностей	n.s.	HL=+0.44 ***, $r\approx 0.35$	HL=+0.35 ***, $r\approx 0.34$

Продолжение таблицы 3

Индикатор	СОЦ-РП	МЕД-РП	МЕД-МП
Лексическая когезия			
Косинусное сходство пар слов	$\Delta=+0.012^{***}$ , $r\approx 0.39$	n.s.	$\Delta=+0.004^*$ , $r\approx 0.24$
Косинусное сходство соседних предложений	$\Delta=+0.045^{***}$ , $r\approx 0.41$	n.s.	n.s.
Грамматическая когезия			
Эксплицитное название	HL=-81 **, $r\approx 0.24$	n.s.	HL=+51.9 **, $r\approx 0.24$
Дискурсивные маркеры локальной связности			
ДМ введения новой информации	HL=+0.14 **, $r\approx 0.26$	HL=-0.15 **, $r\approx 0.23$	n.s.
ДМ противопоставления	HL=-0.12 *, $r\approx 0.21$	HL=-0.19 ***, $r\approx 0.35$	n.s.
ДМ введения примеров	n.s.	HL=-0.084 ***, $r\approx 0.38$	HL=-0.073 ***, $r\approx 0.35$
ДМ причинно-следственных отношений	n.s.	HL=-0.076 *, $r\approx 0.17$	n.s.
ДМ целевых отношений	n.s.	HL=-0.22 ***, $r\approx 0.28$	HL=-0.73 ***, $r\approx 0.70$
Дискурсивные маркеры глобальной связности			
ДМ порядка следования информации	n.s.	n.s.	HL=-0.134 *, $r\approx 0.22$
ДМ выводов и заключения	n.s.	n.s.	HL=+0.05 *, $r\approx 0.21$

По показателю *косинусного сходства пар слов* и *соседних предложений* в СОЦ-РП фиксируется снижение, что противоречит гипотезе об *экспликации* и говорит о меньшей семантической связанности используемой лексики и меньшей плавности смысловых переходов между предложениями в сравнении с непереводаемыми социологическими статьями. В МЕД-МП наблюдается некоторое снижение *косинусного сходства пар слов*, что указывает на небольшое снижение семантической связанности используемой в МП лексики.

В СОЦ-РП растёт показатель *эксплицитного названия*, что, вероятно, указывает на более активное использование личных и притяжательных местоимений для идентификации именованных сущностей. В МЕД-МП данный показатель, напротив, снижается. При этом доля личных и притяжательных местоимений в МЕД-МП оказалась выше, чем в непереводаемом подкорпусе медицинских статей, в связи с чем можно предположить, что в исходных англоязычных статьях по медицине именованные сущности используются в целом реже, чем в непереводаемых статьях, что и обуславливает полученный результат. В МЕД-РП и МЕД-МП снижается *средняя длина именованных сущностей* и растёт доля *однословных именованных сущностей*, что противоречит гипотезе об *экспликации* и, скорее, отражает особенности употребления имен собственных в оригинальных англоязычных статьях по медицине.

Перейдем к результатам анализа частоты использования разных типов дискурсивных маркеров. Начнем с маркеров *локальной связности*. Маркеры *введения новой информации* демонстрируют противоположную динамику в

подкорпусах ручных переводов. В СОЦ-РП их частота снижается, что противоречит гипотезе об *экспликации* и говорит о несколько менее последовательном введении новой информации в сравнении с непереводаемыми русскоязычными статьями. В то же время в МЕД-РП частота данного типа маркеров растет, что указывает на более явное обозначение переходов между высказываниями по сравнению с непереводаемыми текстами. В СОЦ-РП и МЕД-РП фиксируется рост частоты маркеров *противопоставления*, что может объясняться как *интерференцией* исходных английских текстов, в которых контраст между различными положениями, как правило, выражается более эксплицитно, так и *экспликацией*, проявляющейся в более явном маркировании имплицитно содержащихся в исходных текстах отношений противопоставления. В подкорпусах МЕД-РП и МЕД-МП фиксируется рост частоты маркеров *введения примеров*, что, вероятно, является следствием *интерференции* исходных англоязычных текстов. Некоторый рост маркеров *причинно-следственных отношений* в МЕД-РП согласуется с гипотезой об *экспликации*, поскольку этот тип маркеров позволяет сделать аргументацию более последовательной и логически выстроенной. В МЕД-РП и в особенности МЕД-МП наблюдается рост частоты маркеров *целевых отношений*. В случае с МЕД-РП данный результат может объясняться как *экспликацией* целевых отношений оригинала, так и *интерференцией* английских целевых конструкций. В то же время в случае с МЕД-МП столь стремительный рост частоты маркеров данного типа и столь крупный эффект ( $r \approx 0.70$ ) с большой вероятностью свидетельствуют о калькировании целевых конструкций исходных текстов, что требует особого внимания при постредактировании. Что касается маркеров *глобальной связности*, то в МЕД-МП наблюдается некоторое увеличение частоты маркеров, обеспечивающих *порядок следования информации*, и некоторое снижение частоты маркеров *выводов и заключения*. В случае с МП, где отсутствует намеренное вмешательство человека, такие различия между подкорпусами отражают особенности организации англоязычных оригиналов.

Перейдем к характеристике *интерференция*. В качестве иллюстрации в таблице 4 представлены некоторые результаты анализа *частеречных n-грамм*.

Таблица 4 – Индикаторы характеристики *интерференция (n-граммы)*

N-грамма	СОЦ-РП	МЕД-РП	МЕД-МП
NOUN	$\Delta = +1.96$ ***, $r \approx 0.41$	$\Delta = +1.76$ ***, $r \approx 0.43$	$\Delta = +2.94$ ***, $r \approx 0.65$
NOUN NOUN	$\Delta = +0.87$ ***, $r \approx 0.38$	$\Delta = +0.57$ *, $r \approx 0.20$	$\Delta = +1.71$ ***, $r \approx 0.59$
NOUN ADJF NOUN	$\Delta = +0.35$ **, $r \approx 0.30$	$\Delta = +0.27$ **, $r \approx 0.24$	$\Delta = +0.93$ ***, $r \approx 0.69$
VERB	$\Delta = -0.76$ ***, $r \approx 0.49$	n.s.	n.s.
ADVB	$\Delta = -0.28$ **, $r \approx 0.28$	$\Delta = -0.36$ ***, $r \approx 0.38$	n.s.
ADVB VERB	$\Delta = -0.11$ ***, $r \approx 0.39$	n.s.	n.s.
ADJF	$\Delta = +0.80$ *, $r \approx 0.26$	n.s.	$\Delta = +1.44$ ***, $r \approx 0.42$
MDL_VRB	$\Delta = -0.14$ ***, $r \approx 0.40$	$\Delta = -0.21$ ***, $r \approx 0.44$	n.s.
MDL_VRB INFN	$\Delta = -0.10$ ***, $r \approx 0.39$	$\Delta = -0.13$ ***, $r \approx 0.37$	n.s.
MDL WRD	$\Delta = -0.06$ *, $r \approx 0.25$	$\Delta = -0.11$ **, $r \approx 0.33$	$\Delta = +0.09$ ***, $r \approx 0.48$
SBJ PRCL (частица <i>бы</i> )	$\Delta = -0.08$ ***, $r \approx 0.46$	$\Delta = -0.02$ ***, $r \approx 0.36$	n.s.
COPULA	$\Delta = -0.10$ ***, $r \approx 0.36$	$\Delta = -0.08$ **, $r \approx 0.33$	$\Delta = -0.18$ ***, $r \approx 0.60$
AUX PRTS	n.s.	$\Delta = -0.15$ ***, $r \approx 0.31$	$\Delta = -0.42$ ***, $r \approx 0.70$
PRTF	n.s.	$\Delta = -0.27$ ***, $r \approx 0.28$	$\Delta = -0.32$ **, $r \approx 0.31$

Продолжение таблицы 4

N-грамма	СОЦ-РП	МЕД-РП	МЕД-МП
COMMA PRTF	n.s.	$\Delta=-0.13^*$ , $r\approx 0.23$	$\Delta=-0.25^{***}$ , $r\approx 0.42$
PER PRN	$\Delta=-0.42^{***}$ , $r\approx 0.57$	$\Delta=-0.15^{***}$ , $r\approx 0.45$	$\Delta=-0.29^{***}$ , $r\approx 0.72$
PER PRN VERB	$\Delta=-0.25^{***}$ , $r\approx 0.58$	$\Delta=-0.10^{***}$ , $r\approx 0.48$	$\Delta=-0.26^{***}$ , $r\approx 0.77$
DMSTR PRN	$\Delta=-0.37^{***}$ , $r\approx 0.40$	$\Delta=-0.49^{***}$ , $r\approx 0.47$	n.s.
COMMA SCONJ	$\Delta=-0.53^{***}$ , $r\approx 0.63$	$\Delta=-0.25^{***}$ , $r\approx 0.35$	$\Delta=-0.17^{**}$ , $r\approx 0.28$
COMMA PRN CONJ W	$\Delta=-0.18^{***}$ , $r\approx 0.43$	$\Delta=-0.16^{***}$ , $r\approx 0.55$	$\Delta=-0.24^{***}$ , $r\approx 0.71$

Во всех переводных подкорпусах фиксируется снижение доли имен существительных и комбинаций с ними. Результат свидетельствует о снижении степени номинативности переводов, вероятно, вследствие *интерференции*: хотя в обоих языках научные тексты отличаются высокой степенью номинативности, в англоязычных текстах она выражена в меньшей мере. Поскольку высокая степень номинативности является стилеобразующей чертой русского научного стиля, при переводе и постредактировании следует целенаправленно контролировать её, чаще прибегая к номинализации (субстантивации) глаголов там, где это необходимо. В СОЦ-РП снижение номинативности сопровождается ростом доли глаголов и инфинитивов, что говорит об увеличении степени глагольности ручных переводов по социологии. Кроме того, в СОЦ-РП наблюдается рост доли наречий и комбинаций наречий с глаголами, сопровождающийся уменьшением доли полных имен прилагательных – результат, связанный с тем, что снижение степени номинативности закономерно влечёт за собой перераспределение зависимых от неё атрибутивно-обстоятельственных отношений. При этом в МЕД-РП наблюдается рост доли наречий, а в МЕД-МП – значительное снижение доли прилагательных.

В СОЦ-РП и МЕД-РП фиксируется рост доли модальных глаголов и их комбинаций с инфинитивом, а также модальных слов. Можно предположить, что данный результат является следствием *интерференции* модальных синтаксических структур английского языка, в котором конструкции с модальными глаголами являются наиболее продуктивным средством выражения деонтической и эпистемической модальности (конструкции типа *must be noted*, *should be considered*; *may require*, *might indicate* и т. д.) и при переводе нередко передаются как модальными глаголами, так и модальными словами русского языка. В МЕД-МП доля модальных слов, напротив, снижается. Интересен рост доли сослагательной частицы *бы* (*SBJ\_PRCL*) в СОЦ-РП и МЕД-РП, который, предположительно, может быть обусловлен *интерференцией* английских конструкций с *would* и *could* в эпистемическом значении, широко используемых в англоязычном научном дискурсе в качестве средств хеджирования и при переводе нередко передающихся сослагательным наклонением (*this would indicate* → *это указывало бы на*; *could be attributed to* → *можно было бы объяснить*).

Во всех переводных подкорпусах растёт доля глаголов-связок, причём в МЕД-МП эффект наибольший ( $r\approx 0.60$ ). Вероятно, этот результат связан с калькированием английских конструкций с *to be*: в английском глагол-связка в составном именном сказуемом обязательна, и, по всей видимости, она нередко переносится в текст перевода даже в случае, когда может быть опущена.

В МЕД-РП и в особенности в МЕД-МП наблюдается рост доли аналитических страдательных конструкций, вероятно, связанный с *интерференцией*: аналитические конструкции английского оригинала (типа *was conducted, were measured*) могут переноситься в перевод в виде соответствующих структур ПЯ (*было проведено, были измерены*), хотя допустимым средством передачи страдательного значения являются возвратные глаголы (*проводилось, измерялись*).

В МЕД-РП и МЕД-МП растет доля комбинаций *COMMA+PRTF* (запятая и полное причастие), что указывает на более частое употребление обособленных причастных оборотов. Результат, вероятно, обусловлен невозможностью использования в английском номинативных рамочных конструкций (конструкции типа *определенные на предыдущем этапе методологические ограничения* в английском языке не употребляются (\* *defined at the previous step methodological limitations*), допустимый в данном случае вариант *methodological limitations defined at the previous step* в переводе с большей вероятностью будет передан причастным оборотом: *методологические ограничения, определенные на предыдущем этапе*).

Во всех переводных подкорпусах растет доля личных местоимений и их комбинаций с личными формами глагола (*PER\_PRN+VERB*). Результат, по всей видимости, обусловлен влиянием англоязычных конвенций научного письма, где использование личных местоимений для выражения личности исследователя более приемлемо и в ряде случаев более предпочтительно, чем в русскоязычной научной традиции, в которой предпочтение часто отдается формам страдательного залога (Ср.: *We found / can see a correlation between the variables* → *Мы обнаружили / можем наблюдать корреляцию между переменными* вместо, например, *Выявлена / наблюдается корреляция между переменными*). В СОЦ-РП и МЕД-РП растет доля указательных местоимений, что, вероятно, связано с *экспликацией*: переводчики стремятся к более явному маркированию референциальных связей между элементами текста с помощью указательных местоимений (*этот, такой* и т. д.).

Примечательно увеличение во всех переводных подкорпусах доли комбинаций *COMMA+SCONJ* (запятая и подчинительный союз) и *COMMA+PRN\_CONJ\_W* (запятая и союзное местоименное слово), по сути вводящих разные типы придаточных предложений. В СОЦ-РП и МЕД-РП это может быть следствием проявления *экспликации* логико-смысловых отношений между частями высказываний или *интерференции* синтаксических структур англоязычных оригиналов. В случае с МП, где отсутствует сознательное намерение сделать перевод более понятным, результат связан с воспроизведением системой МП сложноподчиненных структур исходного текста.

Перейдем к анализу дополнительной группы индикаторов (Табл. 5).

Таблица 5 – Индикаторы, отражающие особенности научного стиля ПЯ

Индикатор	СОЦ-РП	МЕД-РП	МЕД-МП
Доля существительных	$\Delta=+2.74$ ***, $r\approx 0.46$	$\Delta=+2.37$ ***, $r\approx 0.40$	$\Delta=+3.86$ ***, $r\approx 0.59$
Доля сущ. в ед. числе	$\Delta=+4.26$ ***, $r\approx 0.37$	n.s.	$\Delta=+5.67$ ***, $r\approx 0.58$
Доля абстрактных сущ.	$\Delta=+3.19$ **, $r\approx 0.26$	$\Delta=-2.36$ **, $r\approx 0.23$	n.s.
Доля прилагательных	$\Delta=+1.36$ ***, $r\approx 0.36$	$\Delta=+1.01$ **, $r\approx 0.24$	$\Delta=+2.24$ ***, $r\approx 0.49$
Доля кратких прилагательных	$\Delta=-1.65$ ***, $r\approx 0.43$	$\Delta=-1.39$ ***, $r\approx 0.36$	n.s.

Продолжение таблицы 5

Индикатор	СОЦ-РП	МЕД-РП	МЕД-МП
Доля прил. в превосх. степени	n.s.	n.s.	$\Delta=-0.42 *$ , $r\approx 0.27$
Доля глаголов	$\Delta=-1.02 ***$ , $r\approx 0.38$	$\Delta=-0.98 ***$ , $r\approx 0.36$	$\Delta=-0.51 *$ , $r\approx 0.23$
Доля страдат. залога	$\Delta=+2.79 **$ , $r\approx 0.28$	n.s.	$\Delta=-4.12 **$ , $r\approx 0.28$
Доля причастий	$\Delta=+4.21 ***$ , $r\approx 0.43$	n.s.	n.s.
Доля причастий в препозиции	n.s.	n.s.	$\Delta=+3.42 ***$ , $r\approx 0.30$
Доля глаг. будущ. времени	$\Delta=-1.39 **$ , $r\approx 0.36$	$\Delta=-0.60 *$ , $r\approx 0.24$	n.s.
Доля глаг. наст. времени	n.s.	n.s.	$\Delta=+18.75 ***$ , $r\approx 0.53$
Доля глаг. прош. времени	n.s.	n.s.	$\Delta=-18.37 ***$ , $r\approx 0.51$
Доля глаг. несов. вида	n.s.	n.s.	$\Delta=+8.73 ***$ , $r\approx 0.55$
Личные и притяжат. местоимения	$\Delta=-0.68 ***$ , $r\approx 0.41$	$\Delta=-0.19 *$ , $r\approx 0.21$	$\Delta=-0.22 *$ , $r\approx 0.26$
Относительные местоимения	$\Delta=-0.59 ***$ , $r\approx 0.50$	$\Delta=-0.29 ***$ , $r\approx 0.31$	$\Delta=-0.41 ***$ , $r\approx 0.44$
Указательные местоимения	n.s.	$\Delta=-0.66 ***$ , $r\approx 0.48$	n.s.
Определительные местоимения	n.s.	$\Delta=-0.20 ***$ , $r\approx 0.37$	$\Delta=-0.19 ***$ , $r\approx 0.40$
Неопределённые местоимения	n.s.	$\Delta=-0.075 ***$ , $r\approx 0.34$	$\Delta=-0.045 *$ , $r\approx 0.22$
ДМ высокой степени уверенности	$HL=-0.12 ***$ , $r\approx 0.35$	$HL=-0.10 ***$ , $r\approx 0.28$	n.s.
ДМ средней степени уверенности	$HL=-0.19 ***$ , $r\approx 0.36$	$HL=-0.19 ***$ , $r\approx 0.34$	n.s.
ДМ акцентирования внимания	$HL=+0.013 *$ , $r\approx 0.22$	n.s.	n.s.
ДМ апелляции к фоновым знаниям	n.s.	n.s.	$HL=+0.076 ***$ , $r\approx 0.39$

Во всех переводных подкорпусах снижается доля существительных и растёт доля глаголов, что подтверждает вывод о снижении степени номинативности переводов. В СОЦ-РП и МЕД-МП фиксируется снижение доли существительных в форме единственного числа, что может быть следствием калькирования форм множественного числа английских оригиналов, в которых данная форма в целом используются более активно, в частности для обозначения общих понятий или совокупности предметов (*patients, cells*), тогда как в русской научной традиции в этой функции предпочтительна форма единственного числа (*пациент, клетка*). Доля абстрактных существительных снижается в СОЦ-РП, а в МЕД-РП – растёт. В СОЦ-РП такое снижение может быть обусловлено как *интерференцией* (если в оригиналах доля абстрактных существительных уступает доле конкретных), так и снижением номинативности переводов.

Снижение доли прилагательных в переводных подкорпусах коррелирует со снижением доли существительных: прилагательные функционируют преимущественно как определения при существительных. В корпусах СОЦ-РП и

МЕД-РП наблюдается увеличение доли кратких прилагательных, что может быть обусловлено переносом английских составных именных сказуемых с прилагательным (*the indicators are comparable* → *показатели сопоставимы*), однако данное предположение требует верификации на параллельном подкорпусе. В МЕД-МП фиксируется рост доли прилагательных в форме превосходной степени, что в ряде случаев может нарушать нормы русского научного стиля и должно учитываться на этапе постредактирования.

В корпусе СОЦ-РП наблюдается снижение доли страдательных форм глагола, что согласуется с результатами анализа частеречных n-грамм (рост комбинаций *PER\_PRN* и *PER\_PRN+VERB*). Данный результат, вероятно, обусловлен *интерференцией* английского языка, современная норма академического письма которого допускает широкое использование активного залога и местоимений 1-го лица там, где в русскоязычных научных текстах предпочтительны формы страдательного залога (Ср. *In this paper we examine...* и *В статье рассматривается...*; *We collected data from* и *Собраны данные из...*). В МЕД-МП доля страдательных форм заметно растёт, что согласуется с уже описанным ростом комбинаций *AUX+PRTS*.

В СОЦ-РП снижается доля причастий, что коррелирует как с повышением доли комбинаций *СОММА+PRN\_CONJ\_W*, а также с приведенным ниже ростом доли относительных местоимений и может быть интерпретировано как проявление синтаксической *интерференции*: вероятно, переводчики воспроизводят англоязычную модель выражения атрибутивных отношений посредством придаточных определительных, не используя в полной мере потенциал причастных конструкций русского языка. В МЕД-МП снижается доля причастий перед определяемым ими существительным, что согласуется с ростом комбинаций *СОММА+PRTF*, обусловленным невозможностью использования в английском номинативных рамочных конструкций: вероятно, система МП воспроизводит причастные обороты оригинала (*the method proposed by...* → *метод, предложенный (кем-то)*; *the data obtained during...* → *данные, полученные в ходе (чего-то)*), не выбирая варианты с причастием в препозиции (*предложенный (кем-то) метод*; *полученные в ходе (чего-то) данные*).

Интересно, что в СОЦ-РП и МЕД-РП растёт доля глаголов в форме будущего времени. С практической точки зрения данный результат заслуживает внимания при постредактировании, поскольку для русского научного стиля будущее время не столь характерно. В подкорпусе МЕД-МП резко снижается доля глаголов в форме настоящего времени, что сопровождается стремительным увеличением доли глаголов в форме прошедшего времени. Результат, вероятно, обусловлен воспроизведением системой МП временных форм оригинала в тексте перевода, что может нарушать нормы русского научного стиля. Кроме того, в МЕД-МП значительно снижается доля глаголов несовершенного вида. Поскольку категория вида в английском языке отсутствует, системе МП не всегда удастся правильно соотносить глагол оригинала с необходимой видовой категорией в переводе. При постредактировании МП необходима целенаправленная проверка видо-временных значений с заменой форм прошедшего времени на формы настоящего и глаголов совершенного вида на глаголы несовершенного там, где это необходимо.

Рост доли личных и притяжательных, а также относительных местоимений во всех корпусах подтверждает выводы, сделанные на основе анализа частеречных программ. В МЕД-РП растет доли указательных местоимений, что, вероятно, является проявлением *экспликации* референциальных связей в тексте. В МЕД-РП и МЕД-МП наблюдается рост доли определительных и неопределённых местоимений. Можно предположить, что данный результат обусловлен *интерференцией*: поскольку в английском языке у имён существительных присутствует категория определённости / неопределённости, в большинстве случаев она требует явного маркирования с помощью разнообразных определительных элементов (типа *every, each, any, some* и т. п.), которые при переводе могут переноситься в текст в виде соответствующих им местоимений ПЯ (типа *каждый, любой / всякий, некоторый*).

В СОЦ-РП и МЕД-РП фиксируется рост частоты маркеров *высокой и средней степеней уверенности*. Результат согласуется с ростом модальных глаголов и слов в ручных переводах и может отражать влияние богатого арсенала средств эпистемической модальности англоязычного научного дискурса, которые при переводе передаются соответствующими маркерами хеджирования на ПЯ. Снижение маркеров *акцентирования внимания* в СОЦ-РП, вероятно, является следствием *интерференции*: считается, что данный тип маркеров менее характерен для английских научных текстов, чем для русских. Снижение маркеров *апелляции к фоновым знаниям* в МЕД-МП может быть обусловлено тем, что подобные конструкции используются в русскоязычных научных статьях для подтверждения легитимности выдвигаемых положений несколько чаще, чем в англоязычных.

## ЗАКЛЮЧЕНИЕ

В настоящем диссертационном исследовании проведён комплексный корпусный анализ феномена переводческого языка на материале переводов научных статей с английского языка на русский. В ходе работы были рассмотрены существующие подходы к пониманию феномена переводческого языка и его характеристик, а также особенности научного стиля в русском и английском языках, легшие в основу выбора отдельной группы индикаторов анализа этого феномена. Для эмпирического исследования переводческих характеристик были отобраны 46 формальных индикаторов четырёх основных характеристик переводов (*упрощения, нормализации, экспликации, интерференции*), а также индикаторы, отражающие соответствие текста нормам русского научного стиля. Разработан специализированный программный комплекс, обеспечивший автоматизированный сбор и анализ трёх сопоставимых корпусов общим объёмом около 1,27 млн словоформ. Статистическая значимость различий между подкорпусами оценивалась с применением t-критерия Уэлча и U-критерия Манна–Уитни с расчётом величины эффекта  $g$ .

Исследование показало, что характеристики переводческого языка проявляются во всех трёх переводных подкорпусах, однако степень и характер их проявления варьируются в зависимости от предметной области (социологическая / медицинская тематика) и типа перевода (ручной / машинный), причём машинные переводы демонстрируют более выраженные отклонения от норм русского научного стиля, чем ручные. Установлено, что проявления характеристик

неоднозначны с точки зрения их влияния на качество перевода и могут быть разделены на две группы: (а) Проявления, требующие учёта при переводе и постредактировании, поскольку связаны со стилистическими отклонениями от норм русского научного стиля и могут оказывать влияние на качество перевода. Для каждого из проявлений сформулированы практические рекомендации для переводчиков и постредакторов. (б) Проявления, фиксирующие отличие переводов от непереводаемых текстов и либо не влияющие на качество перевода, либо требующие дальнейших исследований для установления характера их влияния.

Таким образом, поставленная цель исследования – установление индикаторов феномена переводческого языка, влияющих на качество переводов научных статей с английского языка на русский – достигнута. Определён набор конкретных индикаторов, фиксирующих стилистические отклонения от норм русского научного стиля и требующих учёта при переводе и постредактировании. Подтверждена гипотеза о наличии статистически значимых различий между переводными и непереводаемыми научными статьями по ряду индикаторов, соответствующих четырём характеристикам переводческого языка. Разработанный программный комплекс обеспечивает возможность автоматизированного выявления рассматриваемых в настоящей работе индикаторов и может быть использован для установления аспектов перевода, требующих особого внимания при переводе и постредактировании.

Перспективы дальнейших исследований связаны с двумя направлениями: (1) с расширением корпусной базы исследования за счёт включения в анализ параллельных корпусов текстов, что позволит проследить, каким именно образом особенности исходных текстов влияют на проявление обнаруженных особенностей переводов; (2) с включением в исследование текстов статей иных предметных областей, а также текстов иных стилей и жанров, что позволит проверить универсальность выявленных закономерностей на новом языковом материале.

## **СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИОННОГО ИССЛЕДОВАНИЯ**

### *Статьи в журналах, рекомендованных ВАК при Министерстве науки и высшего образования РФ:*

1. Серова О. У. Автоматизированное исследование индикаторов переводческого языка на материале переводов научных статей / О. У. Серова // Terra Linguistica. – 2025. – Т. 16. – № 1. – С. 61–81. – DOI: 10.18721/JHSS.16105 (1,31 п. л.).

2. Серова О. У. Количественный анализ индикаторов феномена translationese в переводах медицинских научных статей / О. У. Серова // Известия Российского государственного педагогического университета имени А. И. Герцена. – 2025. – № 217. – С. 312-326. – DOI: 10.33910/1992-6464-2025-217-312-326 (0,94 п. л.).

3. Серова О. У. Феномен translationese в машинном переводе научных статей с английского языка на русский (количественные характеристики) / О. У. Серова // Этнопсихоллингвистика. – 2025. – № 4 (23). – С. 71-88. – DOI: 10.31249/epl/2025.04.05 (1,1 п. л.).

***Статьи в сборниках трудов научных конференций***

4. Серова О. У. Исследование феномена translationese (на материале машинного перевода научных текстов с английского на русский язык) / О. У. Серова // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во С.-Петербур. ун-та. – 2024. – С. 286-293 (0,5 п. л.).